The Role of the Spatial Boundary in Autopoiesis

Nathaniel Virgo, Matthew Egbert and Tom Froese

Centre for Computational Neuroscience and Robotics
University of Sussex, Brighton, UK
nathanielvirgo@gmail.com
matthew.egbert@gmail.com
t.froese@gmail.com

Abstract. We argue that the significance of the spatial boundary in autopoiesis has been overstated. It has the important task of distinguishing a living system as a unity in space but should not be seen as playing the additional role of delimiting the processes that make up the autopoietic system. We demonstrate the relevance of this to a current debate about the compatibility of the extended mind hypothesis with the enactive approach and show that a radically extended interpretation was intended in one of the original works on autopoiesis. Additionally we argue that the definitions of basic terms in the autopoietic literature can and should be made more precise, and we make some progress towards such a goal.

1 Introduction

The idea of autopoiesis is a venerable part of the artificial life tradition. Ideas from the theory of autopoiesis formed part of the foundations on which the field of artificial life was built [1] and have been widely cited ever since.

However, over its lifetime the idea of autopoiesis has meant different and in many cases quite incompatible things to different authors. An important part of the subject's maturation will be to determine more precisely whether these alternative approaches are compatible with each other and what, if anything, forms their common theoretical core (c.f. [5]).

We suggest that much of the conflict in this field comes from the conflation of two concepts that should be kept distinct: the *physical boundary* of an autopoietic system, which is produced by the system and makes an important contribution to the working of the system; and what we call its *operational limits*, which determine which processes are part of the system. The goal of this paper is to make these concepts, and the distinction between them, as clear as possible. Failure to do this in the past has lead to a kind of internalism in which the network of processes that constitute an organism is seen to lie entirely within its physical boundary, an idea that sits uncomfortably with the conception of cognition as relational. We believe that by clarifying this distinction and introducing new terminology for it we will make a direct contribution towards current modelling and theoretical work.

One particular point of relevance for this discussion is a current debate about whether the extended mind hypothesis [2] is compatible with the 'enactive' approach developed by Varela and colleagues [8], in which both autopoiesis and an extended approach (in which cognitive processes can take place outside an organism's physical bounds) play central roles. The possibility of incompatibility between the two was first raised by Wheeler in a talk at last year's Artificial Life conference [9, 10], which has subsequently been the target of a critical analysis by Di Paolo [4].

Wheeler's argument forms a useful point of departure for clarifying the way in which enactive cognitive science conceives of the complex relationship between life and mind, as well as its operational understanding of cognition. It progresses with the following steps:

1. Autopoiesis is a non-negotiable component of enactive cognitive science.
2. Autopoiesis is a type of self-organization defined by the production of a physical boundary that distinguishes the system as a material unity.
3. One interpretation of the primary literature is that "autopoiesis = life = cognition."
4. It seems to follow from steps 2 and 3 that the enactive approach is committed to the claim that the cognitive system is co-extensive with the living system, entailing that both are bounded by the living system's physical membrane.
5. Since cognition is therefore internal to the physical boundary of the autopoietic system, the enactivist cannot endorse the extended mind hypothesis.

Since the autopoietic and enactive traditions have always insisted on the relational nature of cognition as emerging out of the dynamics of a brain-body-world systemic whole, this conclusion might come as a surprise. Is the enactive approach really committed to the claim that cognition is something that happens within the spatial bounds of an organism?

One way to dissolve this particular incompatibility is to admit that the "life = cognition" slogan has outlived its usefulness. This is the approach pursued by Di Paolo [4], who emphasises the non-reducibility (non-intersection) but mutual interdependence of the metabolic (constitutive) and cognitive (relational) domains of discourse in the autopoietic tradition. On this view the enactive approach to cognition is committed to neither an internalist nor an externalist position: "as relational in this strict sense, *cognition has no location*." [4, p. 19, original emphasis].

Though compatible with Di Paolo's argument, our position has a different focus. It is Wheeler's interpretation of autopoiesis as a self-sustaining network of molecular processes that occur *within* a physical boundary (step 2 above) that gives him the original motivation for his argument. This internalist interpretation of autopoiesis may indeed be held by some of the idea's current proponents, but we argue for a different interpretation, in which the physical 'constitutive' processes by which an organism's structure is produced need not all take place within its physical boundary. By spelling out in detail the distinction between the spatial boundary and the operational limits we arrive at a view of enactive

cognitive science that cannot be considered 'internalist' neither on the cognitive nor the physical, metabolic level.

The Changing Definition of Autopoiesis It is important to be clear that autopoiesis is not a well-defined concept, even though it sometimes appears to be. Much of the primary and secondary literature is written in a style that suggests the theory being discussed is fully developed and quite formally defined; but in fact the meaning of many of the key terms, including the word 'autopoiesis' itself, change quite fluidly from publication to publication. For instance, in [6] autopoiesis is explicitly presented as a theory that applies to all life, whether multicellular or unicellular, whereas in [7] and later publications the idea is said to apply only to single cells, with multicellular organisms requiring a special 'second-order' autopoiesis.

In order for the theory to be moved forward it is important to continue to work on these definitions. It is not enough to quote one of the varying definitions that Maturana and Varela gave, since these were neither precise nor unchanging. Moreover they depend on the meanings of words such as 'process' which as far as we are aware were never defined in any of the primary literature. Our approach is to try to reveal some of the key concepts by stripping away some of the convoluted and overly formal language, focusing instead on sharpening our (pre-formal) intuitions.

2 Defining Operational Closure and its Limits

What is a process? Since the word 'process' has such a key role in all of Maturana and Varela's definitions of autopoiesis it is surprising that little seems to be written about its precise meaning (though see [3]). A related under-asked question is what it means for a process to be *enabled by* or *dependent upon* another process. Since our discussion below also hinges heavily on the concept, we will briefly summarise our (somewhat tentative) intuitions about what a definition might look like here.

A tentative definition of a process might be that it is something that happens repeatedly, or which tends to happen whenever the right conditions are met (examples of processes that meet this definition include: the fermentation of sugar into alcohol; diffusion of heat from hot to cold bodies). There are several properties that are shared by every process, at least in the physical/chemical domain: every such process transforms something into something else (a chemical reaction transforms its reactants into its products; a transport process transforms the spatial distribution of a substance; friction transforms kinetic energy into heat). All such processes also have conditions which must be met in order for them to take place, or which affect the rate at which they occur. These can be trivial (e.g. simply the presence of the reactants) or more complex (such as the presence of enzymes and a specific temperature).

Note that on this definition, processes are separate from the dynamics: the dynamics are the ways in which variables change over time, whereas processes

are things that cause them to change. Thus one can model the dynamics of a system without modelling the processes that underly them.

Importantly, the operation of a process can modify the conditions that determine whether another process takes place. Rather than a process B depending directly upon another process A, we have a situation in which process A produces something which is a required condition for process B to occur. Process B then depends on A, via the conditions that process A helps to generate. These relationships of dependence can form networks of processes with the interesting property of operational closure which we shall now discuss.

Operational Limits and Spatial Boundaries In this section we introduce the term *operational limits* to describe which processes should be seen as an operational part of a system. We discuss the relation between operational limits and *spatial boundaries* and show why the two notions are often conflated, despite being quite distinct.

To define the operational limits of a system, it is necessary to understand the notion of *operational closure*. Figure 1 depicts a hypothetical system of processes that are dependent upon or enabled by each other. These interdependencies are depicted in the figure as arrows connecting the processes. Given such a network of processes & relationships of dependence, we can use the following definition to identify operationally closed parts of the network. However, before we give this definition, we think that it is important to point out that the application of this definition requires as a precondition, the identification of all of the processes and relationships of dependence. As we mentioned in the previous section, at this stage, neither 'process' nor 'dependence' (sometimes referred to as 'conditioning' or 'enabling') are well defined. Thus far, researchers have depended upon intuitive understanding of these phenomena, but they are in need of formalization if we are to consider operational closure rigorously defined.

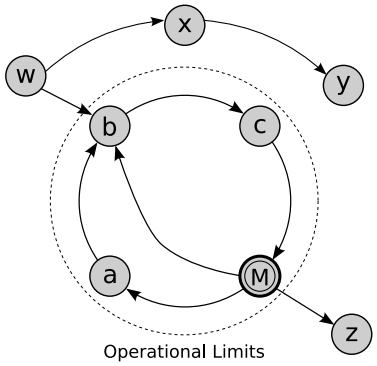

Fig. 1. A hypothetical network of processes connected by interdependencies. Lettered circles represent processes and arrows represent 'enable'. M represents a process that generates the spatial boundary.

Definition Given a collection of processes $\mathcal{C}$, we can identify an operationally closed subset of those processes, $\mathcal{S}$ such that for every constituent process P, the following conditions are true.

1. Another process P' is conditioned by process P
2. Process P is conditioned by another processP''
3. P' and $P'' \in \mathcal{S}$
4. P' and P'' can be (but are not required to be) the same process.

In graph theory terms, this defines a strongly connected subgraph of the directed graph of process dependencies. Assuming that all of the processes and interdependencies are included in figure 1, processes w, x, y and z are not part of any operationally closed network. This is the case because each one of these processes does not depend upon another process which is in turn dependent upon the original process. Take as a case in point, process x which is dependent upon w which is not dependent upon any process in the system. An absence of cyclical dependence indicates an absence of operational closure. In contrast, processes a, b, c and M form an operationally closed network. Process c depends upon b which depends upon a which depends upon M which depends upon c, closing the loop and making the set of four processes operationally closed. A second, smaller operationally closed loop exists, consisting only of processes M, b, and c. These are the only operationally closed loops within this system.

It is not difficult to find examples of processes that have cycles of dependency on a variety of scales from subsystems of an organism to relationships of dependence on an ecological or global scale. Furthermore, operationally closed networks of processes may or may not involve organisms at all. This is not problematic for the theory of autopoiesis as autopoietic systems are the subset of operationally closed systems that produce a spatially bounded structure.

Let us now consider this spatial boundary. There is no requirement that any of the above processes occur inside or outside of the spatial boundary. It may be tempting to think that because processes a, b, c and M are all inside the operational limits, they are also inside the spatial boundary, but this is not necessarily the case. It might be that only processes a, b, and z are inside the spatial boundary. The spatial boundary is not the same as the operational limits.

In a way cell membranes, or spatial boundaries in general, seem similar to the operational limits in that they define a boundary inside which certain processes lie and others do not. For this reason, it has been tempting for some authors to depict spatial boundaries on relational diagrams as a circle surrounding a number of processes, similar to the depiction of the operational limits of an organization. To do this however is to commit the error of conflating operational limits with spatial boundaries. Figure 1 is not drawn in physical space – it is a relational diagram of processes. As such, it is inappropriate to depict boundaries in their spatial form (e.g. as an encircling). It is only appropriate to depict them as yet another process (e.g. process M in our diagram) or set of processes that has various interdependencies with the other processes in the network.

On one hand this error does not seem too serious if, for example, one is drawing informal diagrams intended to get a point across. However, there are a number of serious conceptual errors that can be caused by confusing the spatial and relational domain. For example, processes can span the spatial boundaries of an organism (e.g. ion pumps in cell-membranes or the production of heat by warm-blooded animals affecting the animal's environment). This possibility is lost when spatial boundaries and process relationships are conflated and physical and relational structures are plotted in the same space.

A related error is the inappropriate inflation of the importance of the spatial boundary. The spatial boundary is undoubtedly important in maintaining the conditions necessary for many ongoing processes in living organisms. While these are indeed important contributions, we do not believe that they are of a different *type* of contribution than the other enabling processes that form living organisms. In fact, we believe that the inflation of the importance of the spatial boundary runs contrary to one of the more provocative ideas to come from the autopoietic school of thought. Namely, that the spatial boundary of the organism is not actually equivalent to the limits of the organism – that the organism includes processes that are not occuring within its spatial boundary.

3 Extended Autopoiesis

In this section we defend our claim that the operationally closed network that constitutes an autopoietic unity can include processes that occur outside of its spatial boundary by showing that this was the interpretation intended in one of the earliest pieces of literature on the subject, Maturana and Varela's *Autopoiesis and Cognition* [6].

This does not validate our claim entirely; what matters to science is what is useful to us in the present day, not the precise words that were first written 37 years ago. However the original exposition is quite clear and we hope that by re-examining it with a simple example we can better express our own perspective.

In [6], Maturana and Varela introduce us to the concept of homeostatic machines. These are defined as machines which maintain constant, or within a limited range of values, some of their variables, a definition which will be familiar to most of us. However this definition is followed by an important clarification which we take as fundamental to how the rest of the theory is to be interpreted. Since the clarification of this definition is so important we quote the whole paragraph:

> "There are machines which maintain constant, or within a limited range of values, some of their variables. The way this is expressed in the organization of these machines must be such as to define the process as occurring completely within the boundaries of the machine which the very same organization specifies. Such machines are homeostatic machines and all feedback is internal to them. If one says that there is a machine M, in which there is a feedback loop through the environment so that the effects of its output affect its input, one is in fact talking about a larger machine M' which includes the environment and the feedback loop in its defining organization." [6, section 1.2.a]

This can be clarified with an example. Let us consider a mechanical thermostat. This is an archetypal example of a homeostatic machine (though of course it is not autopoietic). The variable which it keeps within bounds is the temperature of a room. However, according to the quoted paragraph it is not correct (in the autopoietic language) to think of the thermostat as being the box on the wall that is connected to a heater and contains a thermocouple, because

this machine (machine M) has a feedback loop that runs through the environment. When the temperature drops, the thermocouple breaks a connection, which causes the heater (not part of machine M) to be switched off, causing the temperature to drop again. Since the thermostat relies on this feedback loop for its operation, we should actually define the thermostat as a larger machine (machine M') which includes the heater, the air in the room, and the feedback loop that passes through them.

Why is this so important? The above quoted paragraph is positioned directly before the definition of an autopoietic machine is spelled out[1], and just below that we are given the following key statement:

> "Therefore, an autopoietic machine is an homeostatic (or rather relation-static) system which has its own organization (defining network of relations) as the fundamental variable which it maintains constant." [*ibid.*]

Autopoietic systems, then, are to be seen as homeostatic machines. It follows that their definition must be expanded in the same way if they rely on a feedback loop that runs through their environment.

Wheeler [9] gave the example of an earthworm, which builds tunnels held open by a sticky secretion that helps to digest its food. We can try to see the worm as an autopoietic system (and hence an homeostatic system) whose operational limits are defined by its physical boundary (its skin). However, the worm relies on the effects of its secretions; this is a feedback loop which runs through its environment. The above quoted paragraph from [6] thus compels us to redefine the system so that it includes not only the worm itself but also the secretions and their effects. The autopoietic system that constitutes the worm is not coextensive with the unity that we refer to as "the worm," it is much bigger. This will be the case for most if not all organisms, since most organisms rely not only on sensory-motor loops that run through their environment but also on nutrients that are recycled externally to them.

4 Future Considerations

In the past, because it is easily visualizable, because terminology was confusing, and perhaps because of some difficulties associated with translation, we have

[1] The full definition of autopoiesis as given in [6] is as follows. Note that this version of the definition hinges on the constitution of a concrete unity in space but does not specify that this unity must be bounded by a distinct membrane.

> An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network.

seen the physical boundary of an autopoietic system as playing a special role in both the physical and the relational domains. But here we have argued that in the relational domain the spatial boundary should take its place among the other enabling conditions. In the physical domain it plays an important role in helping to define the organism as a distinct unity but it plays no special role in the relational domain, except perhaps in that it enables a great number of processes.

We have shown some of the implications of this for the debate about the compatibility between autopoiesis and the extended mind hypothesis, and we believe that it is relevant to much current theoretical and modelling work.

We have also begun working towards precise definitions of some basic concepts in the autopoietic theory, which were previously absent. The definitions we have outlined are tentative. However we have tried to express them in such a way that the intended interpretation is clear. We hope that future authors will try to give similarly precise definitions of basic terms. The theory of autopoiesis can only become stronger as a result.

Acknowledgements

This paper arose from discussions at the Life and Mind seminars at Sussex, and on the associated blog (http://lifeandmind.wordpress.com). The contribution of Ezequiel Di Paolo to this discussion was particularly influential.

References

1. P Bourgine and F. J. Varela. Introduction: Towards a practice of autonomous systems. In F. J. Varela and P. Bourgine, editors, *Proceedings of the First European Conference on Artificial Life*. MIT Press, 1992.
2. A. Clark and D. Chalmers. The extended mind. *Analysis*, 58:10–23, 1998.
3. E. A. Di Paolo. Artificial life and historical processes. In J. Kelemen and P. Sosik, editors, *Proceedings of ECAL 2001*, pages 649–658, Prague, 2001. Springer-Verlag.
4. E. A. Di Paolo. Extended life. *Topoi*, 28(1):9–12, 2009.
5. J. Kiverstein and A. Clark. Introduction: Mind embodied, embedded, enacted: One church or many? *Topoi*, 28(1):1–7, 2009.
6. H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. Kluwer Academic Publishers, Dordrecht, Holland, 1980.
7. H. R. Maturana and F. J. Varela. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala Publications, Boston, MA, 1987.
8. E. Thompson. *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. The Belknap press of Harvard University Press, Cambridge, MA, 2007.
9. M. Wheeler. Autopoiesis, enactivism, and the extended mind (abstract). In Bullock. S, J. Noble, R. Watson, and M. A. Bedau, editors, *Artificial Life XI: Proc. of the 11th Int. Conf. on the Simulation and Synthesis of Living Systems*, page 819, Cambridge, MA, 2008. MIT Press.
10. M. Wheeler. Minds, things and materiality. In C. Renfew and L. Malafouris, editors, *The Cognitive Life of Things: Recasting the Boundaries of the Mind*. McDonald Institute for Archaeological Research Publications, Cambridge, in press.